ON A CLASS OF MOMENT PROBLEMS

J. H. B. KEMPERMAN
UNIVERSITY OF ROCHESTER

Introduction

Consider a class $\mathcal{M}$ of probability measures on a measurable space X and measurable functions g_j and h on X. In a typical moment problem we want further information on the possible sets of values taken by the integrals $\mu(g_j)$ and $\mu(h)$ as μ runs through $\mathcal{M}$. And the main purpose of the present paper is to develop into more systematic methods certain principles which in special cases have been found effective for handling such moment problems.

In Sections 2 through 4 we take up certain frequently occurring moment problems where the class $\mathcal{M}$ happens to be convex. In Section 2 the space X can be any locally compact Hausdorff space. For $\{h_j, j \in J\}$ as an arbitrary collection (finite or infinite) of lower semicontinuous functions on X, we establish a condition which is both necessary and sufficient for the existence of a regular probability measure μ on X satisfying $\mu(h_j) \leqq \eta_j$ for all $j \in J$. However, we do assume as a side condition that the h_j dominate each other at infinity in a certain weak sense. This domination condition is void when X is compact and nearly so when $h_j \geqq 0$.

In Sections 3 and 4 we are interested in the smallest value $L(y)$ of $\mu(h)$ when it is known that $\mu \in \mathcal{M}$ and that $\mu(g_j) = y_j$ for $j = 1, \cdots, n$; the space X can be any measurable space. Provided this smallest value $L(y)$ is in fact assumed, it turns out that in the determination of $L(y)$ we only need to consider so called *admissible* measures.

These are defined as the measures $\mu \in \mathcal{M}$ which attain the smallest possible value $\mu(\psi)$ for some linear combination ψ of the form $\psi = h - d_1 g_1 - \cdots - d_n g_n$. In the special case that $\mathcal{M}$ consists of all probability measures on X, we have admissibility if and only if the measure is carried by the set of minima of some such linear combination ψ.

In Sections 5 and 6 we are interested in bounds for and inequalities between the different moments of a sum $S_n = Z_1 + \cdots + Z_n$ of independent random variables Z_i. Here, the Z_i may have different distributions subject to certain restrictions on these distributions. The resulting collection $\mathcal{M}$ of possible distributions of S_n is usually not convex.

An essential use is made of the fact that each cumulant $\kappa_j(S_n)$ of S_n is equal to the sum of the $\kappa_j(Z_i)$. The set $K[q]$ of possible q-tuples $(\kappa_1(Z), \cdots, \kappa_q(Z))$ is usually not a convex subset of R^q. It turns out that for large n the existing

This research was partially supported by the National Science Foundation.

inequalities between the different moments $E(S_n^j)$, $j = 1, \cdots, q$, are more or less determined by the structure of the convex hull of $K[q]$.

In Section 5 the resulting inequalities are worked out in detail for the case that $0 \leqq Z \leqq 1$ and $q \leqq 4$. Section 6 contains among other things an explicit method for determining the best possible upper bound on $E(\exp\{tS_n\})$ subject only to the condition that $0 \leqq Z_i \leqq 1, E(Z_i^j) = c_j$ for $j = 1, \cdots, m$, all i, while $E(S_n^j) = d_j$ for $j = m+1, \cdots, q$. Here $q \leqq 2m + 1$ and the Z_i may have different distributions.

2. A general moment problem

2.1. In the present section we treat a frequently occurring moment problem which may involve infinitely many side conditions. In the sequel, X denotes a locally compact Hausdorff space made into a measurable space by the σ-field of all Borel subsets of X. We shall often employ lower semicontinuous functions $h: X \to R$. This means that $h(x) \leqq \liminf_{y\to x} h(y)$ for all $x \in X$; equivalently, the set $\{x: h(x) > c\}$ is always open; such a function h is bounded below on each compact subset of X.

Further on, $\{h_j, j \in J\}$ will denote a given finite or infinite collection of lower semicontinuous functions on X (such as the characteristic functions of the open subsets of X). This collection is sometimes denoted by $\mathscr{H}$. Next, $\{\eta_j, j \in J\}$ denotes a given real valued function on the index set J. Finally, $\mathscr{M}_*$ will stand for the class of all regular probability measures μ on X such that each h_j is integrable relative to μ in such a way that

$$\mu(h_j) = \int h_j(x)\mu(dx) \leqq \eta_j \qquad \text{for each } j \in J; \tag{2.1}$$

$\mu(A) \leqq \eta_j$ if h_j is the characteristic function of the open set A.

We shall be interested in establishing sufficient conditions for $\mathscr{M}_*$ to be nonempty. We may expect that such a result would enable us to handle many other moment problems simply by adjoining new functions h_j to the system $\mathscr{H}$ and adding new conditions of the type (2.1), (see for instance [5], p. 569). Also observe that (2.1) allows us to formulate a condition of the form

$$\int g_i(x)\mu(dx) = \rho_i, \tag{2.2}$$

where g_i is a *continuous* function on X and i may run through an index set I of any cardinality. All one needs to do is to adjoin both g_i and $-g_i$ to the system $\mathscr{H}$.

Let R^J denote the collection of all real valued functions $\beta(\cdot)$ on J such that $\beta(j) = \beta_j = 0$ for all but finitely many $j \in J$. Similarly, let R_+^J denote the collection of all nonnegative functions in R^J. It is clear from (2.1) that

$$\int [\alpha_0 + \sum_J \beta_j h_j]\, d\mu \leqq \alpha_0 + \sum_J \beta_j \eta_j, \tag{2.3}$$

as soon as $\alpha_0 \in R$, $\beta(\cdot) \in R_+^J$ and $\mu \in \mathscr{M}_*$. Therefore, in order that $\mathscr{M}_*$ be nonempty it is at least necessary that

$$\alpha_0 + \sum_J \beta_j h_j \geqq 0, \qquad \alpha_0 \in R,\ \beta \in R_+^J \Rightarrow \alpha_0 + \sum_J \beta_j \eta_j \geqq 0. \tag{2.4}$$

Here, if $\phi\colon X \to R$ then $\phi \geqq 0$ denotes that $\phi(x) \geqq 0$ for all $x \in X$.

REMARK 2.1. Actually, (2.4) is already a necessary condition for a much weaker property than $\mathscr{M}_*$ being nonempty. Namely, suppose instead that for each choice of the finite subset J' of J and each choice of the numbers $\delta_j > 0$, $j \in J'$, there exists a probability measure μ on X such that

$$\int h_j\, d\mu \leqq \eta_j + \delta_j \qquad \text{for each } j \in J'. \tag{2.5}$$

Obviously this implies (2.4).

2.2. The following examples will show that in fact condition (2.4) is *not sufficient* for $\mathscr{M}_*$ to be nonempty.

Let $X = R$ and let $\mathscr{M}_*$ be determined by the conditions

$$\int x^2 \mu(dx) \geqq 1, \qquad \int e^{-x^2} \mu(dx) = 1. \tag{2.6}$$

Then $\mathscr{M}_*$ is clearly empty though (2.4) is satisfied. More precisely, one may take in this case $J = \{1, 2\}$, $h_1(x) = 1 - x^2$ with $\eta_1 = 0$ and $h_2(x) = 1 - e^{-x^2}$ with $\eta_2 = 0$. That (2.4) is satisfied follows, for instance, from Remark 2.1 by replacing $\eta_2 = 0$ by $\eta_2 = \delta$ with $\delta > 0$ arbitrarily small, and observing that there do exist probability measures μ on R for which $\int x^2\, d\mu \geqq 1$ and $\int e^{-x^2}\, d\mu \geqq 1 - \delta$.

As a second counterexample, take X as the discrete space $X = \{1, 2, 3, \cdots\}$. Let $\{h_j, j \in J\}$ and $\{\eta_j, j \in J\}$ be such that $h_j(x) \geqq 0$ always while

$$\eta_j \geqq \limsup_{x \to \infty} h_j(x), \qquad j \in J, \tag{2.7}$$

and further

$$\inf_j \{\eta_j / h_j(x) : j \in J,\ h_j(x) > 0\} = 0 \quad \text{for all } x \in X. \tag{2.8}$$

(For instance, one may take $h_j(x) = x^{-j}$ and $\eta_j = 1/j!$, where $j = 1, 2, \cdots$; or take $\eta_j = 0 = \lim_{x\to\infty} h_j(x)$ and $\sup_j h_j(x) > 0$; or take $h_j(x) = j + j^3/x$ and $\eta_j = j^2$, where $j = 1, 2, \cdots$.)

Condition (2.4) is an immediate consequence of (2.7). On the other hand, let μ be any nonnegative measure on X satisfying (2.1). Then $\eta_j \geqq h_j(x)\mu(\{x\})$ for all $j \in J$, $x \in X$ and we conclude from (2.8) that $\mu(\{x\}) = 0$ for all $x \in X$, so that μ cannot possibly be a probability measure. In other words, $\mathscr{M}_*$ is empty.

As a last counterexample, take again $X = \{1, 2, 3, \cdots\}$ and let (2.1) be of the form $\mu(g_i) = 0$ for all $i \in I$. Here, we take $g_i\colon X \to R$ such that $g_i(x) \to 0$ as $x \to \infty$, for all $i \in I$, so that condition (2.4) is trivially satisfied. Finally, suppose that for each $x \in X$ and each $\varepsilon > 0$ there exists an index $i \in I$ such that $g_i(x) > 0$

and $g_i(x') \geqq -\varepsilon g_i(x)$ for all $x' \in X$; (it would be sufficient that a single function g_i be positive and nonincreasing toward 0). If μ were a probability measure satisfying $\mu(g_i) = 0$, we would have

$$0 = \mu(g_i) \geqq -(1 - \mu(\{x\}))\varepsilon g_i(x) + \mu(\{x\})g_i(x), \tag{2.9}$$

implying that $\mu(\{x\}) \leqq \varepsilon/(1 + \varepsilon)$. But $x \in X$ and $\varepsilon > 0$ are arbitrary, thus, $\mu = 0$ so that $\mathscr{M}_*$ is in fact empty.

REMARK 2.2. We can think of other necessary conditions besides (2.4) for $\mathscr{M}_*$ to be nonempty. For instance, one would be

$$h_j(x) > 0 \quad \text{for all } x \in X \Rightarrow \eta_j > 0. \tag{2.10}$$

In view of Fatou's lemma, another necessary condition would be

$$h_j(x) \geqq 0, \quad \beta_j \geqq 0, \qquad \sum_j \beta_j h_j(x) = \infty \quad \text{for all } x \Rightarrow \sum_j \beta_j \eta_j = \infty, \tag{2.11}$$

where j is to be restricted to some denumerable subset of J. In the case of conditions of the form $\mu(g_i) = \rho_i, i = 1, 2, \cdots$, the dominated convergence theorem yields as a further necessary condition that $\Sigma\, \alpha_i \rho_i = 0$ as soon as $\Sigma\, \alpha_i g_i(x) = 0$ for all $x \in X$ and further that $\Sigma |\alpha_i g_i(x)| \leqq \Sigma\, \beta_j h_j(x)$ for some choice of the numbers $\beta_j \geqq 0$ and the functions $h_j \geqq 0$ in $\mathscr{H}$ with $\Sigma\, \beta_j \eta_j < \infty$, j being restricted to a countable subset of J.

Some of the above additional necessary conditions are in fact violated by the counterexamples outlined in this section. Nevertheless, if possible we would clearly prefer to avoid using conditions of the type (2.11) since they are hard to verify.

One would also like to keep the system $\mathscr{H} = \{h_j, j \in J\}$ as small as possible so that (2.4) may not be too hard to verify. Naturally, using the properties of an integral (such as Fatou's lemma and linearity) one can usually enlarge $\mathscr{H}$ considerably without affecting the class $\mathscr{M}_*$, but we will refrain from doing this.

DEFINITION 2.1. *If h and ϕ are real valued functions on the locally compact space X, we will say that h is* dominated below at infinity *by ϕ when, for each number $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset X$ such that*

$$h(x) \geqq -\varepsilon|\phi(x)| \quad \textit{for each } x \notin K_\varepsilon. \tag{2.12}$$

Observe that it would be sufficient that h be nonnegative or that X itself be compact.

If (2.12) is replaced by $h(x) \leqq \varepsilon|\phi(x)|$ for each $x \notin K_\varepsilon$, we will say that h is dominated above at infinity by ϕ. If both properties hold we say that h is dominated at infinity by ϕ.

DEFINITION 2.2. *Let Φ denote the class of all* nonnegative *functions $\phi : X \to R$ which admit at least one representation as*

$$\phi(x) = \alpha_0 + \sum_J \beta_j h_j(x), \qquad x \in X, \tag{2.13}$$

with $\alpha_0 \in R$, $\beta(\cdot) \in R_+^J$. Observe that each $\phi \in \Phi$ is lower semicontinuous and also that Φ is a convex cone in the obvious sense.

Only for the purpose of a proof, we further associate to each $\phi \in \Phi$ a nonnegative number $q(\phi)$ such that $q(\phi) \geqq \alpha_0 + \Sigma_J \beta_j \eta_j$ for at least one representation of the form (2.13). Observe that, by (2.1), we have $\mu(\phi) \leqq q(\phi)$ for each $\mu \in \mathscr{M}_*$.

In the sequel we shall make a frequent use of the function g_0 on X defined by $g_0(x) = 1$ for all $x \in X$. Therefore, $\mu(g_0) = 1$ for each $\mu \in \mathscr{M}_*$.

THEOREM 2.1. *Suppose that, for each $j \in J$, the function h_j is dominated below at infinity by some $\phi_j \in \Phi$; it would be sufficient that $h_j \geqq 0$. Suppose further that the special function g_0 is dominated at infinity by some $\phi_0 \in \Phi$, and let $\mathscr{M}_*$ denote the collection of all regular probability measures on X satisfying* (2.1). *Then*

(i) *the collection $\mathscr{M}_*$ is nonempty if and only if condition* (2.4) *holds;*

(ii) *$\mathscr{M}_*$ is a convex set which is compact in the weak* topology.*

Here, as usual, the weak* topology is taken relative to the class $\mathscr{K}(X)$ of all continuous functions f on X having a compact support (that is, $\{x : f(x) \neq 0\}$ has a compact closure). Thus a net $\{\mu_i, i \in I\}$ of regular Borel measures converges to μ if and only if $\mu_i(f) \to \mu(f)$ for all $f \in \mathscr{K}(X)$. Concrete applications of Theorem 2.1 may be found in [5].

For the moment, consider a pair h_j, ϕ_j as in the theorem. Since $\phi_j \geqq 0$ it is integrable relative to a nonnegative finite measure μ as soon as $\mu(\phi_j) < \infty$. We claim that this implies that h_j is at least improperly integrable so that $\mu(h_j)$ is well defined. After all, $h_j(x) \geqq -\varepsilon\phi_j(x)$ for x outside some compact set K_ε, while on K_ε the lower semicontinuous function h_j is bounded below.

Assertion (i) of Theorem 2.1 was already established in [5] pp. 565 and 570. (Apply Theorem 4.1 of [5] with F as the linear manifold spanned by g_0 and the h_j, $j \in J$, and take F^+ as the convex cone of all $f \in F$ such that $f \geqq \psi$ for some (upper semicontinuous) function ψ of the form $\psi = q(\phi)g_0 - \phi$; here $\phi \in \Phi$, while the scalars $q(\phi)$ are chosen such that $\mu \in \mathscr{M}_*$ if and only if $\mu(g_0) = 1$ and $\mu(\psi) \geqq 0$ for all such ψ; the present condition (2.4) corresponds to the condition $-g_0 \notin F^+$ of [5]; condition (4.1) of [5] appears unnecessary.)

The proof in [5] used the classical Hahn-Banach theorem together with the Riesz representation theorem. The proof below of (i) and (ii) relies more heavily on the linear space $\mathscr{M}(X)$ of all real valued finite signed regular measures on X, made into a locally convex topological vector space by means of the weak* topology. Further,

$$B(X) = \{\mu \in \mathscr{M}(X) : \mu \geqq 0, \|\mu\| \leqq 1\}, \tag{2.14}$$

will denote the set of all nonnegative measures $\mu \in \mathscr{M}(X)$ of total mass $\leqq 1$. It is essential for the proof that $B(X)$ is not only convex but also compact (in the weak* topology).

PROOF OF THEOREM 2.1. In the sequel, we shall assume all the conditions of Theorem 2.1 and further condition (2.4), since, otherwise, $\mathscr{M}_*$ would be empty.

For each *finite* subset J' of J, let $\mathscr{M}_*(J')$ denote the set of all $\mu \in B(X)$ satisfying

$$\mu(g_0) = 1, \qquad \mu(h_j) \leqq \eta_j \quad \text{if} \quad j \in J', \tag{2.15}$$

and

$$\mu(\phi_0) \leqq q(\phi_0), \qquad \mu(\phi_j) \leqq q(\phi_j) \quad \text{if} \quad j \in J'. \tag{2.16}$$

We easily verify that the class $\mathscr{M}_*$ is precisely equal to the intersection of the collection of all such classes $\mathscr{M}_*(J')$. Moreover, this collection has the finite intersection property, since $\mathscr{M}_*(J') \cap \mathscr{M}_*(J'') = \mathscr{M}_*(J' \cup J'')$. Hence, in order to prove the theorem, that is, in order to prove that $\mathscr{M}_*$ is nonempty and compact it suffices to prove that each individual class $\mathscr{M}_*(J')$ is nonempty and compact.

From now on, let J' be fixed. For convenience, we shall take $J' = \{1, 2, \cdots, n\}$. Using condition (2.4) and the definitions of Φ and $q(\phi)$, we easily see that for any choice of the real constants α, $\beta_j \geqq 0, j = 1, \cdots, n$, and $\gamma_j \geqq 0, j = 0, 1, \cdots, n$, we have

$$\alpha + \sum_1^n \beta_j \eta_j + \sum_0^n \gamma_j q(\phi_j) \geqq 0, \tag{2.17}$$

as soon as

$$\alpha g_0(x) + \sum_1^n \beta_j h_j(x) + \sum_0^n \gamma_j \phi_j(x) \geqq 0 \quad \text{for all } x \in X. \tag{2.18}$$

Introducing $\phi_{n+j} = h_j, j = 1, \cdots, n$, and

$$\zeta_j = q(\phi_j), \quad j = 0, 1, \cdots, n, \quad \zeta_{n+j} = \eta_j, \quad j = 1, \cdots, n, \tag{2.19}$$

this implication can be restated as

$$\alpha + \sum_{j=0}^{2n} \gamma_j \phi_j \geqq 0, \quad \gamma_j \geqq 0 \Rightarrow \alpha + \sum_{j=0}^{2n} \gamma_j \zeta_j \geqq 0. \tag{2.20}$$

Here, the functions $\phi_j, j = 0, 1, \cdots, 2n$, are all lower semicontinuous. Moreover, $\phi_j \geqq 0, j = 0, 1, \cdots, n$; thus $\zeta_j \geqq 0, j = 0, \cdots, n$. Moreover, $g_0 \equiv 1$ is dominated at infinity by ϕ_0, while ϕ_{n+j} is dominated below at infinity by $\phi_j, j = 1, \cdots, n$. Finally, $\mathscr{M}'_* = \mathscr{M}_*(J')$ can now also be described as the collection of all $\mu \in B(X)$ such that

$$\mu(g_0) = 1, \qquad \mu(\phi_j) \leqq \zeta_j, \qquad j = 0, 1, \cdots, 2n. \tag{2.21}$$

We shall first prove the following three results:

(i) for $j = 0, 1, \cdots, n$, the functions $\mu \to \mu(\phi_j)$ are lower semicontinuous on $B(X)$; in particular we have that the set $\{\mu \in B(X) \colon \mu(\phi_j) \leqq c\}$ is always closed, $j = 0, 1, \cdots, n$;

(ii) on the set $A(c) = \{\mu \in B(X) \colon \mu(\phi_0) \leqq c\}$ (with c as a finite constant) the function $\mu \to \mu(g_0)$ is continuous; thus $\{\mu \in B(X) \colon \mu(\phi_0) \leqq c, \mu(g_0) = 1\}$ is a closed set;

(iii) let $1 \leqq j \leqq n$; then on the set

$$B_j(c) = \{\mu \in B(X) \colon \mu(g_0) = 1, \mu(\phi_j) \leqq c\}, \tag{2.22}$$

(with c as a finite constant) the function $\mu \to \mu(\phi_{n+j}) = \phi(h_j)$ is lower semicontinuous.

One easily verifies that (i), (ii), (iii) together imply that the set $\mathscr{M}'_*$ defined by (2.21) is in fact a closed subset of $B(X)$ and therefore compact.

To prove (i) let $0 \leqq j \leqq n$ be fixed. Then ϕ_j is nonnegative and lower semicontinuous, in which case

$$\mu(\phi_j) = \sup \{\mu(f): f \in \mathscr{K}(X), 0 \leqq f \leqq \phi_j\}, \tag{2.23}$$

holds for each $\mu \in B(X)$ (see [1], p. 104). Here, by the definition of the weak* topology each function $\mu \to \mu(f)$ is a continuous function of μ. Hence, it follows from (2.23) that the function $\mu \to \mu(\phi_j)$ is lower semicontinuous on $B(X)$.

For the proof of (ii), we use that g_0 is dominated at infinity by $\phi_0 \geqq 0$. Let $\varepsilon > 0$ and choose the compact set K_ε such that $1 = |g_0(x)| \leqq \varepsilon\phi_0(x)$ for each $x \notin K_\varepsilon$. Next choose ψ_ε in $\mathscr{K}(X)$ such that $\psi_\varepsilon(x) = 1 = g_0(x)$ for $x \in K_\varepsilon$ and $0 \leqq \psi_\varepsilon(x) \leqq 1$, otherwise. It follows that

$$|\mu(g_0) - \mu(\psi_\varepsilon)| \leqq \int_{K_\varepsilon'} d\mu \leqq \varepsilon \int \phi_0(x)\, d\mu \leqq \varepsilon c, \tag{2.24}$$

as long as $\mu \in A(c)$. Hence, on $A(c)$ the function $\mu \to \mu(g_0)$ is the uniform limit of the continuous functions $\mu \to \mu(\psi_\varepsilon)$ and therefore itself continuous.

To prove (iii), let $1 \leqq j \leqq n$ be fixed. We know that the function $\phi_{n+j} = h_j$ on X is dominated below at infinity by the nonnegative function ϕ_j. Hence, for each $\varepsilon > 0$ there exists a compact set K_ε in X such that

$$h_j(x) + \varepsilon\phi_j(x) \geqq 0 \quad \text{for } x \notin K_\varepsilon. \tag{2.25}$$

Here, the left side defines a lower semicontinuous function which therefore is bounded below on K_ε. Consequently, there exists a constant a_ε such that $a_\varepsilon g_0 + h_j + \varepsilon\phi_j \geqq 0$ everywhere. In view of a relation of the type (2.23), the function $\mu \to \mu(a_\varepsilon g_0 + h_j + \varepsilon\phi_j)$ is lower semicontinuous throughout $B(X)$. Hence, on $B_j(c)$ the function

$$\mu(h_j + \varepsilon\phi_j) = \mu(a_\varepsilon g_0 + h_j + \varepsilon\phi_j) - a_\varepsilon \tag{2.26}$$

is lower semicontinuous. But on $B_j(c)$ we also have

$$|\mu(h_j) - \mu(h_j + \varepsilon\phi_j)| \leqq \varepsilon\mu(\phi_j) \leqq \varepsilon c. \tag{2.27}$$

Therefore, on $B_j(c)$ the function $\mu \to \mu(h_j)$ is the uniform limit of lower semicontinuous functions and thus itself lower semicontinuous.

It remains to prove that the compact set $\mathscr{M}'_*$ is nonempty. Let D denote the set of all points $w \in R^{2n+1}$ such that there exists a measure $\mu \in B(X)$ satisfying

$$\mu(g_0) = 1, \qquad \mu(\phi_j) \leqq w_j, \qquad j = 0, 1, \cdots, 2n. \tag{2.28}$$

Clearly the set D is convex. Using the above results (i), (ii), (iii) and the fact that $B(X)$ itself is compact, we easily see that the set D is also closed. Thus, D is equal to an intersection of a collection of closed half spaces.

It is given that (2.20) holds and we must prove that the set $\mathscr{M}'_*$ defined by (2.21) is nonempty. In other words, we must prove that (2.20) implies $z \in D$,

where $z = (\zeta_0, \zeta_1, \cdots, \zeta_{2n})$. It suffices to prove that $z \in H$, where H is any closed half space containing D. Let this half space have the form

$$H = \left\{w : \alpha + \sum_{j=0}^{2n} \gamma_j w_j \geqq 0\right\}, \tag{2.29}$$

where α and γ_j denote real constants. Since $D \subset H$ and D is unbounded in each positive w_j direction we must have that $\gamma_j \geqq 0$.

Considering a probability measure ε_x on X supported by a single point $x \in X$, we see that $w_j = \phi_j(x), j = 0, 1, \cdots, 2n$, defines a point $w \in D \subset H$; hence, we have that

$$\alpha + \sum_{j=0}^{2n} \gamma_j \phi_j(x) \geqq 0 \quad \text{for all } x \in X. \tag{2.30}$$

Invoking (2.20), we conclude that $\alpha + \Sigma_{j=0}^{2n} \gamma_i \zeta_j \geqq 0$, that is, $z \in H$. This completes the proof of Theorem 2.1.

REMARK 2.3. In Theorem 2.1 the condition (2.4) can also be replaced by

$$\alpha_0 + \sum_J \beta_j h_j > 0, \quad \alpha_0 \in R, \quad \beta \in R_+^J \Rightarrow \alpha_0 + \sum_J \beta_j \eta_j > 0, \tag{2.31}$$

or by

$$\alpha_0 + \sum_J \beta_j h_j > 0, \quad \alpha_0 \in R, \quad \beta \in R_+^J \Rightarrow \alpha_0 + \sum_J \beta_j \eta_j \geqq 0. \tag{2.32}$$

After all, given the other (domination type) conditions of the theorem, we have the implications

$$(2.4) \Rightarrow \mathscr{M}_* \text{ nonempty} \Rightarrow (2.31) \Rightarrow (2.32) \Rightarrow (2.4). \tag{2.33}$$

Here, the first implication follows from Theorem 2.1, while the others are more or less obvious. For an important special case (with X compact) the equivalence between $\mathscr{M}_* \neq 0$ and (2.31) is due to Ky Fan ([3], p. 68).

The following variation of Theorem 2.1 is often useful in applications.

THEOREM 2.2. *As in Theorem* 2.1, *assume that* $g_0 \equiv 1$ *is dominated at infinity by some* ϕ_0 *in* Φ *and further that each* $h_j, j \in J$, *is dominated below at infinity by some* ϕ_j *in* Φ. *Assume also that* $\mathscr{M}_*$ *is nonempty; equivalently, assume that* (2.4) *holds.*

Next, let f *be a fixed upper semicontinuous function on* X *which is dominated above at infinity by some* ϕ *in* Φ, *and define*

$$q(f) = \inf \{\alpha + \sum \beta_j \eta_j : \alpha + \sum \beta_j h_j \geqq f\}; \tag{2.34}$$

here, α *ranges through* R *while* $\beta(\cdot)$ *ranges through* R_+^J.

Clearly, $M(f) \leqq q(f) < \infty$, *where*

$$M(f) = \sup \{\mu(f) : \mu \in \mathscr{M}_*\}. \tag{2.35}$$

We assert that in fact $M(f) = q(f)$ *and further that the supremum in* (2.35) *is assumed. In fact, the set of* $\mu \in \mathscr{M}_*$ *satisfying* $\mu(f) = M(f)$ *is nonempty, convex and compact (in the weak* topology).*

PROOF. Let f be dominated at infinity by $\tilde{\phi} \in \Phi$; in particular $f \leqq \phi$ for some $\phi \in \Phi$, thus, $q(f) < \infty$. Put $\tilde{h} = -f$ so that $\tilde{h}$ is lower semicontinuous and dominated below by the function $\tilde{\phi}$ in Φ. Adjoining $\tilde{h}$ to $\mathscr{H} = \{h_j, j \in J\}$, one obtains a new class

(2.36) $$\mathscr{M}_*(\gamma) = \{\mu \in \mathscr{M}_*: \mu(\tilde{h}) \leqq -\gamma\} = \{\mu \in \mathscr{M}_*: \mu(f) \geqq \gamma\},$$

which depends on the choice of γ. It follows from Theorem 2.1 that $\mathscr{M}_*(\gamma)$ is always compact and convex. From (2.35), it is empty when $\gamma > M(f)$ and nonempty when $\gamma < M(f)$. Letting $\gamma \uparrow M(f)$, we conclude that $\{\mu \in \mathscr{M}_*: \mu(f) = M(f)\}$ is a nonempty compact and convex set.

Theorem 2.1 also supplies a necessary and sufficient condition for $\mathscr{M}_*(\gamma)$ to be nonempty, namely, condition (2.4) applied to $\mathscr{H} \cup \{\tilde{h}\}$ instead of $\mathscr{H}$. It turns out that $\mathscr{M}_*(\gamma)$ is nonempty if and only if $\gamma \leqq q(f)$. Consequently, we have $M(f) = q(f)$.

REMARK 2.4. The condition of Theorem 2.2 that f be dominated at infinity by some $\phi \in \Phi$ cannot be omitted. For instance, take $X = [1, +\infty)$ and $\{h_j, j \in J\}$ as the single function $h(x) = x^2 + x^{-2}$. Thus $\mathscr{M}_*$ consists of all $\mu \in B(X)$ with $\mu(g_0) = 1$ and $\mu(h) \leqq \eta$. Finally, let $f(x) = x^2$. Clearly, $M(f) = q(f) = \eta$, but, nevertheless, we have $\mu(f) < \eta$ for all $\mu \in \mathscr{M}_*$.

REMARK 2.5. The assertion $M(f) = q(f)$ of Theorem 2.2 is obviously related to the so called fundamental theorem of linear programming (see [5], pp. 558, 561). Of special interest would be the case where not only the supremum $M(f)$ is attained by at least one measure $\mu_0 \in \mathscr{M}_*$, but where also the infimum $q(f)$ is attained by a pair $\alpha \in R$, $\beta(\cdot) \in R^J_+$. This situation will be taken up in Section 3. We easily verify that the measure μ_0 must be carried by the measurable set S of points $x \in X$ for which $\alpha + \Sigma \beta_j h_j(x) = f(x)$; moreover, $\mu_0(h_j) = \eta_j$ whenever $\beta_j > 0$. Conversely, every $\mu_0 \in \mathscr{M}_*$ with these properties does attain $M(f)$.

3. Admissible measures

In this section, X denotes an arbitrary measurable space and $\mathscr{M}$ a given nonempty *convex* collection of (nonnegative) measures on X. We shall be interested in the best lower bound

(3.1) $$L(y) = L(y \mid h) = \inf \{\mu(h): \mu \in \mathscr{M}, \mu(g) = y\}.$$

Here, $g = (g_1, \cdots, g_n)$ is a given measurable function $g: X \to R^n$ which is integrable relative to each $\mu \in \mathscr{M}$. Further, $h: X \to R$ denotes a given measurable function which is integrable relative to each $\mu \in \mathscr{M}$. Finally, $y = (y_1, \cdots, y_n)$ denotes a variable point in R^n.

Clearly, $L(y) < +\infty$ if and only if y belongs to the so called moment space

(3.2) $$M = \{y \in R^n: \mu(g) = y \text{ for some } \mu \in \mathscr{M}\}.$$

Here, M is a convex subset of R^n since the collection $\mathscr{M}$ was assumed to be convex. We may (and shall) assume that M has a nonempty interior; for, otherwise the components g_j of g would be linearly dependent as far as the measures

$\mu \in \mathscr{M}$ are concerned so that part of the information $\mu(g) = y$ would be redundant.

The function $L(y)$ is clearly convex. Excluding the situation that $L(y) = -\infty$ for all $y \in \text{int}(M)$, it follows that $L(y)$ is finite everywhere on M and even continuous throughout int (M).

For each measurable function $\psi: X \to R$, let us introduce

$$\mu_{\min}(\psi = \inf \{\mu(\psi): \mu \in \mathscr{M}\}; \tag{3.3}$$

put $\mu_{\min}(\psi) = -\infty$ if ψ is not even improperly integrable relative to some $\mu \in \mathscr{M}$. If $\mu_{\min}(\psi)$ is finite, then a measure μ on X will be said to be *critical* relative to ψ if

$$\mu \in \mathscr{M}, \qquad \mu(\psi) = \mu_{\min}(\psi). \tag{3.4}$$

EXAMPLE 3.1. Let $\mathscr{M}$ consist of all probability measures on X. In this case

$$\mu_{\min}(\psi) = \inf \psi = \inf \{\psi(x): x \in X\}. \tag{3.5}$$

Moreover, $\mu_0 \in \mathscr{M}$ is critical relative to ψ if and only if it is carried by the "contact" set

$$S(\psi) = \{x \in X: \psi(x) = \inf \psi\}. \tag{3.6}$$

EXAMPLE 3.2. Let λ be a fixed measure on X and let $0 \leqq a(x) \leqq b(x)$ be given measurable functions on X which are integrable relative to λ. Finally, let $\mathscr{M}$ consist of all measures on X of the form

$$\mu(A) = \int_A \rho(x)\lambda(dx), \qquad a(x) \leqq \rho(x) \leqq b(x). \tag{3.7}$$

This measure μ will be critical relative to a function ψ if and only if the corresponding function $\rho(x)$ is such that $\rho(x) = a(x)$ for almost $[\lambda]$ all x with $\psi(x) > 0$ and further $\rho(x) = b(x)$ for almost $[\lambda]$ all x with $\psi(x) < 0$. Here, we are assuming that $\int |\psi| b \, d\lambda < \infty$.

EXAMPLE 3.3. Let $\mathscr{M}$ consist of all measures on X of the form

$$\mu(A) = \int P(u, A)\nu(du). \tag{3.8}$$

Here, ν denotes an arbitrary probability measure on a fixed measurable space U, while P is a given Markov kernel function of $u \in U, A \subset X$. If $\psi \geqq 0$ is measurable then $\mu(\psi) = \nu(\tilde{\psi})$, where

$$\tilde{\psi}(u) = \int \psi(x)P(u, dx). \tag{3.9}$$

Thus, μ is critical relative to ψ if and only if the corresponding measure ν is carried by the contact set $S(\tilde{\psi}) \subset U$.

LEMMA 3.1. *Consider any function (a so called polynomial) of the special form*

$$\psi(x) = h(x) - \sum_{j=1}^{n} d_j g_j(x), \tag{3.10}$$

where the d_j denote real constants. Then

$$L(z) \geqq \mu_{\min}(\psi) + \sum_{j=1}^{n} d_j z_j \quad \text{for all } z \in M. \tag{3.11}$$

Moreover, given $y \in \operatorname{int}(M)$, one can always choose this polynomial ψ in such a way that

$$L(y) = \mu_{\min}(\psi) + \sum_{j=1}^{n} d_j y_j; \tag{3.12}$$

here, ψ is unique for almost all $y \in \operatorname{int}(M)$.

DEFINITION 3.1. *A measure μ_0 on X will be said to be* admissible *if $\mu_0 \in \mathscr{M}$ and further μ_0 is critical relative to* some *polynomial ψ of the form* (3.10).

For instance, in Example 3.1 a probability measure μ_0 is admissible if and only if it is supported by the contact set $S(\psi)$ of some polynomial ψ.

THEOREM 3.1. *Each admissible measure μ_0 assumes $L(y)$ in the sense that*

$$L(y) = \mu_0(h), \text{ where } y = \mu_0(g). \tag{3.13}$$

Conversely, if $y \in \operatorname{int}(M)$ and $L(y)$ is assumed by $\mu_0 \in \mathscr{M}$, then μ_0 is admissible (and the corresponding polynomial ψ is unique for almost all $y \in \operatorname{int}(M)$).

Consequently, if $L(y)$ is assumed for all $y \in \operatorname{int}(M)$, then (3.13) *with μ_0 running through all admissible measures will yield a parametric representation of the function $L(y)$ at least for $y \in \operatorname{int}(M)$.*

PROOF OF LEMMA 3.1. (Another proof is given in [5], p. 574.) Consider a measure $\mu \in \mathscr{M}$ with $\mu(g) = z$. Integrating (3.10), we find that

$$\mu(h) = \mu(\psi) + \sum_{j=1}^{n} d_j z_j \geqq \mu_{\min}(\psi) + \sum_{j=1}^{n} d_j z_j. \tag{3.14}$$

This implies (3.11). Considering $\mu \in \mathscr{M}$ with $\mu(\psi)$ close to $\mu_{\min}(\psi)$ and then taking $z = \mu(g)$, we see that the constant term $\mu_{\min}(\psi)$ in (3.11) cannot be improved. That is, $\gamma = \mu_{\min}(\psi) + d_1 z_1 + \cdots + d_n z_n$ is the best supporting hyperplane in the direction $(d_1, \cdots, d_n)$ to the convex set Q in R^{n+1} consisting of all points (z, γ) with $z \in M$ and $\gamma \geqq L(z)$.

Conversely, consider a fixed point $y \in \operatorname{int}(M)$. Then through the boundary point $(y, L(y))$ of Q there passes a supporting hyperplane to Q. Since $y \in \operatorname{int}(M)$, this hyperplane is nonvertical and of the form $\gamma = d_0 + \Sigma_{j=1}^n d_j z_j$. That is, $L(z) \geqq d_0 + \Sigma_{j=1}^n d_j z_j$ for all $z \in M$, while $L(y) = d_0 + \Sigma_{j=1}^n d_j y_j$. It follows from the above remarks that necessarily $d_0 = \mu_{\min}(\psi)$, where ψ denotes the polynomial defined by (3.10). This yields assertion (3.12).

The uniqueness of ψ, for almost all $y \in \operatorname{int}(M)$, follows from the well-known uniqueness of a supporting hyperplane through the boundary point $(y, L(y))$ of the convex body Q, again for almost all boundary points, that is, for almost all $y \in \operatorname{int}(M)$.

PROOF OF THEOREM 3.1. Let $\mu_0 \in \mathscr{M}$ be admissible, thus, $\mu_0(\psi) = \mu_{\min}(\psi)$ for some polynomial ψ of the form (3.10). Letting $\mu_0(g) = y$, we have

$$L(y) \leqq \mu_0(h) = \mu_0(\psi) + \sum_{j=1}^{n} d_j y_j = \mu_{\min}(\psi) + \sum_{j=1}^{n} d_j y_j. \tag{3.15}$$

In view of (3.11), the equality sign must hold here, proving the first assertion. From Lemma 3.1, we have for almost all y that this can happen for at most one polynomial ψ.

Conversely, let $y \in \mathrm{int}\,(M)$ be fixed and suppose that $L(y)$ is assumed by $\mu_0 \in \mathscr{M}$; thus, $\mu_0(g) = y$ and $\mu_0(h) = L(y)$. Next, choose the polynomial ψ in such a way that (3.12) holds. Then

$$\mu_0(\psi) = \mu_0(h) - \sum_{j=1}^{n} d_j\mu_0(g_j) = L(y) - \sum_{j=1}^{n} d_j y_j = \mu_{\min}(\psi). \tag{3.16}$$

This shows that μ_0 is critical for ψ so that μ_0 is admissible.

In view of Theorem 3.1, we would like to have applicable sufficient conditions on $\mathscr{M}$, $g = (g_1, \cdots, g_n)$ and h in order that the infimum $L(y)$ be assumed. In Theorem 3.2, we take $\mathscr{M}$ as the collection $\mathscr{M}_*$ described in Theorem 2.1. Thus, adopting the notations and assumptions of Theorem 2.1, $\mathscr{M}_*$ is the class of probability measures on the locally compact space X satisfying (2.1), with $\mathscr{H} = \{h_j, j \in J\}$ as a system of lower semicontinuous functions on X satisfying a certain domination type of condition (which is void when X is compact).

THEOREM 3.2. *Let $\mathscr{M} = \mathscr{M}_*$ be as in Theorem 2.1. Let further $g_j, j = 1, \cdots, n$, be given continuous functions on X each dominated at infinity by some $\phi \in \Phi$ and let $h: X \to R$ be a lower semicontinuous function which is dominated below at infinity by some $\phi \in \Phi$.*

Define $L(y)$ as in (3.1) *and M as in* (3.2). *We assert that $L(y)$ is assumed for each $y \in M$ in the sense that for each $y \in M$ there exists $\mu_0 \in \mathscr{M}_*$ with $\mu_0(g) = y$ and $\mu_0(h) = L(y)$.*

PROOF. Simply adjoin the (lower semicontinuous) functions g_j and $-g_j$ to the given system $\mathscr{H} = \{h_j, j \in J\}$ and take $+y_j$ and $-y_j$ as the corresponding η values. Now the assertion that $L(y)$ is assumed immediately follows from Theorem 2.2 applied to this enlarged system and with $f = -h$.

The following result follows directly from Theorem 3.2 by observing that for a compact space X all domination conditions are void. On the other hand, it would not be very hard to prove Corollary 3.1 directly, namely, by using the simple result (i) used in the proof of Theorem 2.1.

COROLLARY 3.1. *Let X be a compact space and $\{h_j, j \in J\}$ any collection of lower semicontinuous functions on X. Take $\mathscr{M}$ as the collection of all regular probability measures μ on X with*

$$\int_X h_j(x)\mu(dx) \leqq \eta_j \quad \text{for each } j \in J. \tag{3.17}$$

Here, the η_j denote given real numbers. We assume that $\mathscr{M}$ is nonempty. Finally, let $g_j: X \to R$ be continuous, $j = 1, \cdots, n$, and let $h: X \to R$ be lower semicontinuous.

Under these assumptions the infimum $L(y)$ in (3.1) *is assumed for each $y \in M$ so that Theorem* 3.1 *becomes applicable.*

4. Applications

4.1. In the present section we outline just one set of applications of the results in Section 3. We shall take the measurable space X as a compact space and $\mathcal{M}$ simply as the collection of all regular probability measures on X. We further assume that the functions $g_1, \cdots, g_n$ are *continuous*, while for the moment we allow for h any measurable function $h: X \to R$ which is bounded below.

Consider the finite valued function

$$\tilde{h}(x) = \liminf_{y \to x} h(y); \tag{4.1}$$

thus, $\tilde{h}(x) \leqq h(x)$. In fact, $\tilde{h}$ is precisely the largest lower semicontinuous function satisfying $\tilde{h} \leqq h$. We assert that

$$L(y \mid h) = L(y \mid \tilde{h}) \quad \text{for each } y \in \text{int}\,(M). \tag{4.2}$$

One way of seeing this would be to apply Lemma 3.1. Using (3.5), this yields that

$$L(y|h) = \sup_d \left[\sum_{j=1}^{n} d_j y_j + \inf_x \left\{ h(x) - \sum_{j=1}^{n} d_j g_j(x) \right\} \right], \tag{4.3}$$

for each $y \in \text{int}\,(M)$. Here, d runs through all n-tuples $d = (d_1, \cdots, d_n) \in R^n$. Now observe that, since the g_j are continuous, the infimum in (4.3) remains unchanged when h is replaced by $\tilde{h}$, hence, (4.2) obtains.

As a more intuitive proof, let $\varepsilon > 0$ be given and choose the neighborhood U of $y \in \text{int}\,(M)$ such that $L(z|h) > L(y|h) - \varepsilon$ for all $z \in U$. Next, choose $\mu_0 \in \mathcal{M}$ such that $\mu_0(g) = y$ and $\mu_0(\tilde{h}) < L(y|\tilde{h}) + \varepsilon$. We may assume (see [6], p. 95) that μ_0 has a finite support (consisting of at most $n + 2$ points). By a slight movement of these support points we obtain a probability measure μ_1 such that $\mu_1(h) < \mu_0(\tilde{h}) + \varepsilon < L(y|h) + 2\varepsilon$, while $z = \mu_1(g)$ still satisfies $z \in U$. We conclude that

$$L(y|h) - \varepsilon < L(z|h) < L(y|\tilde{h}) + 2\varepsilon. \tag{4.4}$$

This in turn yields (4.2).

4.2. From now on, we shall assume that h itself is lower semicontinuous. As indicated by equation (4.2), this is no real loss of generality (when we want to compute $L(y) = L(y|h)$ for $y \in \text{int}\,(M)$; if $y \in M$ is on the boundary of M, we should replace X by an appropriate compact subset so as to get back at the situation $y \in \text{int}\,(M)$; this can always be done, (see [6], p. 102)).

Knowing that h is lower semicontinuous, we have from Corollary 3.1 (or from an easy direct proof) that $L(y)$ is assumed for each $y \in M$. From now on, in this section, let us restrict y to int (M).

Then we conclude from Theorem 3.2 that the computation of $L(y)$ can be reduced to a study of admissible measures.

In the present case, we have from (3.5) and (3.6) that an admissible measure is any probability measure supported by the contact set $S(\psi)$ of some "polynomial" of the form

$$\psi(x) = h(x) - \sum_{j=1}^{n} d_j g_j(x). \tag{4.5}$$

Letting $d_0 = \inf \psi$, we have

$$S(\psi) = \left\{x \in X : d_0 + \sum_{j=1}^{n} d_j g_j(x) = h(x)\right\}. \tag{4.6}$$

Here, d_0 is such that

$$d_0 + \sum_{j=1}^{n} d_j g_j(x) \leqq h(x) \quad \text{for all } x \in X, \tag{4.7}$$

and d_0 is maximal. Since ψ is lower semicontinuous on the compact space X this contact set $S(\psi)$ is always compact and nonempty. We now conclude from Theorem 3.1 that:

(i) all one needs to do in computing $L(y) = L(y|h)$ for given y is to select the admissible measure μ_0 in such a way that $\mu_0(g) = y$; afterwards, $L(y) = \mu_0(h)$;

(ii) this can always be done, that is, μ_0 can always be found;

(iii) call a polynomial ψ associated to y when $S(\psi)$ carries a probability measure μ_0 with $\mu_0(g) = y$. Such an associated polynomial always exists and almost all y have exactly one associated polynomial. Finally, if $\psi = h - \Sigma_{j=1}^{n} d_j g_j$ is associated to y then $L(y)$ is also given by

$$L(y) = d_0 + \sum_{j=1}^{n} d_j y_j, \qquad d_0 = \inf \psi. \tag{4.8}$$

The reader may enjoy using this principle in solving the following problem. Namely, let Z be a real random variable with $0 \leqq Z \leqq 1$ and the first three moments $E(Z^j) = y_j, j = 1, 2, 3$, given. Let further $0 < \alpha < \beta < 1$ be given numbers. Now determine the best possible upper and lower bounds on $Pr(\alpha < Z < \beta)$. For instance, as to the lower bound $L(y)$, either $L(y) = 0$ or the admissible measure corresponding to y has one of the supports $\{\alpha, \xi_1, \beta, 1\}$, $\{0, \alpha, \xi_2, \beta\}$, $\{\alpha, u, 1\}$, $\{\alpha, v, \beta\}$, $\{0, w, \beta\}$. Here, ξ_1 and ξ_2 are fixed numbers, while u, v, w are variable such that $\alpha < u < \xi_1$, $\xi_1 < v < \xi_2$, $\xi_2 < w < \beta$. More or less the same result holds when $E(g_j(Z)) = y_j$, $j = 1, 2, 3$, and $\{g_0, g_1, g_2, g_3\}$ is a Chebyshev system on $[0, 1]$, $g_0 \equiv 1$.

In this and other applications, the main advantage of the present approach comes from the fact that often a set $S(\psi)$ of the type (4.6), (4.7) must be quite small in some sense. This happens for instance when X is an analytic manifold and the g_j and h are analytic or piecewise analytic. These aspects and further applications will be taken up in a subsequent paper.

5. The moments of a sum

5.1. In the present section, we shall be concerned with the following situation where $\mathscr{M}$ is definitely not convex. Namely, let $X = R^k$ and take $\mathscr{M}$ to be of the form

$$\mathscr{M} = \mathscr{M}^{(n)} = \mathscr{M}_1 * \cdots * \mathscr{M}_n. \tag{5.1}$$

Here, the star denotes convolution while the $\mathscr{M}_i$, $i = 1, \cdots, n$, denote given collections of probability measures on R^k. In other words, $\mathscr{M}^{(n)}$ consists of all convolutions of the form $\mu = \mu_1 * \cdots * \mu_n$, where $\mu_i \in \mathscr{M}_i$, $i = 1, \cdots, n$. It is useful to interpret $\mathscr{M}^{(n)}$ also as the collection of all distributions $\mu(A) = Pr(S_n \in A)$ of so called admissible sums $S_n = Z_1 + \cdots + Z_n$ of n independent random variables $Z_i \in R^k$ with the property that the ith component Z_i has a distribution $\mu_i \in \mathscr{M}_i$, $i = 1, \cdots, n$.

Let q be a fixed positive integer and assume that each measure $\mu \in \mathscr{M}_1 \cup \cdots \cup \mathscr{M}_n$ has all moments $y_j = \int x^j \, d\mu$ of order $|j| \leqq q$. Here, $j = (j_1, \cdots, j_k)$ denotes a multi-index with components $j_r \in Z_+ = \{0, 1, 2, \cdots\}$; further $|j| = \Sigma j_r$.

The moment space M_i corresponding to $\mathscr{M}_i$ will be defined as

$$M_i = \left\{ y \in R^{q^*} : \text{there exists } \mu \in \mathscr{M}_i \text{ with } \int x^j \, d\mu = y_j \quad \text{for all } j \in J \right\}, \tag{5.2}$$

$i = 1, \cdots, n$. Here, J denotes the set of all multi-indices j with $1 \leqq |j| \leqq q$ and q^* their number ($q^* = q$ if $k = 1$). Further, a point $y \in R^{q^*}$ is regarded as having coordinates y_j with j running through J.

Similarly, let $M^{(n)}$ denote the moment space corresponding to $\mathscr{M}^{(n)}$. Thus, $M^{(n)}$ is also the set of points $y = (y_j, j \in J)$ in R^{q^*} such that there exists at least one admissible sum $S_n = Z_1 + \cdots + Z_n$ with $E(S_n^j) = y_j$ for all $j \in J$.

We shall be interested in the different relations between these moments $E(S_n^j)$, $j \in J$, that is, in the structure of $M^{(n)}$. In many applications the class $\mathscr{M}_i$ and thus the moment space M_i, $i = 1, \cdots, n$, are convex, while nevertheless $M^{(n)}$ itself and thus $\mathscr{M}^{(n)}$ are nonconvex.

5.2. In studying $M^{(n)}$ it is only natural to use cumulants κ_j, $j \in J$, since these have the addition property

$$\kappa_j(S_n) = \kappa_j(Z_1) + \cdots + \kappa_j(Z_n) \quad \text{for all } j \in J. \tag{5.3}$$

Let $K_i = K_i[q]$ denote the cumulant space corresponding to $\mathscr{M}_i$ and M_i which consists of all points $z \in R^{q^*}$ such that, for some $\mu \in \mathscr{M}_i$, we have $\kappa_j(\mu) = z_j$ for all $j \in J$. Let $K^{(n)} = K^{(n)}[q]$ denote the analogous cumulant space corresponding to $\mathscr{M}^{(n)}$ and $M^{(n)}$. In other words, $z \in K^{(n)}$ if and only if we can find an admissible sum S_n with $\kappa_j(S_n) = z_j$ for all $j \in J$. It follows from (5.3) that

$$K^{(n)} = K_1 + \cdots + K_n, \tag{5.4}$$

where the addition on the right side is ordinary addition of subsets of the additive group R^{q^*}; thus, $A + B = \{z : z = a + b \text{ for some } a \in A, b \in B\}$.

From now on, let us restrict ourselves to the special case where $\mathscr{M}_1 = \mathscr{M}_2 = \cdots = \mathscr{M}_n = \mathscr{M}$, say. Let M and K denote the moment space and cumulant space, respectively, corresponding to the given class $\mathscr{M}$. It follows from (5.4) that

$$K^{(n)} = K + \cdots + K = K^n, \tag{5.5}$$

(in an obvious notation). We further have the important relations

$$K \subset \frac{1}{n} K^{(n)} = \frac{1}{n} K^n \subset \operatorname{conv}(K). \tag{5.6}$$

Here, $(1/n)K^n = \{z : nz \in K^n\}$ may thus also be regarded as the set of all points z in R^{q^*} of the form

$$z = \left(\frac{1}{n}\kappa_j(S_n), j \in J\right), \tag{5.7}$$

for some admissible sum S_n. Moreover, z belongs to the smaller set K in the chain (5.6) if and only if (5.7) holds for some sum $S_n = Z_1 + \cdots + Z_n$ having independent and *identically distributed* components Z_i.

The following result due to Emerson and Greenleaf ([2], p. 180) will play an important role.

LEMMA 5.1. *Let K be a bounded subset of some Euclidean space and suppose that, for some integer $p \geqq 1$, the set $(1/p)K^p$ has a nonempty interior relative to the minimal flat $\mathscr{L}(K)$ containing K.*

Then there exists a constant $c > 0$ depending on K only such that for any $z \in \operatorname{conv}(K)$ and any positive integer n we have either $z \in (1/n)K^n$ or z has a distance $\leqq c/n$ from the complement of $\operatorname{conv}(K)$, (*taken relative to* $\mathscr{L}(K)$).

5.3. Consider the situation where $\mathscr{M}$ and thus K are fixed while n is large. By (5.6), we always have $K^{(n)} \subset n \operatorname{conv}(K)$. It follows from Lemma 5.1 that in a certain sense $n \operatorname{conv}(K)$ is even a very good approximation to $K^{(n)}$. For instance, under the conditions of the lemma we have that $(1/n)K^{(n)}$ tends to $\operatorname{conv}(K)$ in the Hausdorff metric.

Thus, there are several good reasons for trying to determine $\operatorname{conv}(K)$. The only situation which we shall study in some more detail is that where $k = 1$ and

$$\mathscr{M} = \{\text{all probability measures on } [0, c]\}. \tag{5.8}$$

Here, c denotes a fixed positive constant. In other words, we shall be concerned with the moment space $M^{(n)} = M^{(n)}[q]$ and the cumulant space $K^{(n)} = K^{(n)}[q]$ corresponding to the set of all sums $S_n = Z_1 + \cdots + Z_n$ of real valued independent random variables Z_i with possibly different distributions, but such that $0 \leqq Z_i \leqq c$, $i = 1, \cdots, n$. Also recall that $\kappa_1(S) = E(S) = m$, say; $\kappa_2(S) = E(S - m)^2 = \operatorname{Var}(S)$; $\kappa_3(S) = E(S - m)^3$, while $\kappa_4(S) = E(S - m)^4 - 3[E(S - m)^2]^2$.

5.4. It is well known ([4], p. 106) that the moment space $M[q]$ corresponding to $\mathscr{M}$ is precisely the convex hull of the curve

$$\{y : y = (t, t^2, \cdots, t^q) \text{ for some } 0 \leqq t \leqq c\}. \tag{5.9}$$

Moreover, $y = (y_1, \cdots, y_q) \in R^q$ belongs to $M[q]$ if and only if $\Sigma_0^q a_j y_j \geqq 0$ for every polynomial $\Sigma_0^q a_j x^j$ which is nonnegative on $[0, c]$; here, and in the sequel, $y_0 = 1$. The latter condition can be replaced by a small number of polynomial inequalities in $y_1, \cdots, y_q$, involving so called Hankel determinants.

For low values of q these conditions are as follows. First $0 \leqq y_1 \leqq c$ if $q \geqq 1$; moreover, $y_1^2 \leqq y_2 \leqq cy_1$ if $q \geqq 2$; moreover,

$$\begin{vmatrix} y_1 & y_2 \\ y_2 & y_3 \end{vmatrix} \geqq 0, \qquad \begin{vmatrix} c - y_1 & cy_1 - y_2 \\ cy_1 - y_2 & cy_2 - y_3 \end{vmatrix} \geqq 0 \tag{5.10}$$

if $q \geqq 3$; moreover,

$$\begin{vmatrix} 1 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{vmatrix} \geqq 0, \qquad \begin{vmatrix} cy_1 - y_2 & cy_2 - y_3 \\ cy_2 - y_3 & cy_3 - y_4 \end{vmatrix} \geqq 0 \tag{5.11}$$

if $q \geqq 4$, and so on.

Next, we may write

$$K[q] = UM[q], \qquad M^{(n)}[q] = U^{-1}(K[q]^n), \tag{5.12}$$

where U denotes the usual one to one transformation of moment points $y = (y_1, \cdots, y_q)$ into cumulant points $\kappa = Uy = (\kappa_1, \cdots, \kappa_q)$. This transformation is defined by the formal power series identity

$$\sum_1^\infty \left(\frac{\kappa_j}{j!}\right) w^j = \log\left[1 + \sum_1^\infty \left(\frac{y_j}{j!}\right) w^j\right]; \tag{5.13}$$

hence,

$$\kappa_j = (Uy)_j = \sum (-1)^h \frac{h!\,j!}{r_1! \cdots r_j!} \left(\frac{y_1}{1!}\right)^{r_1} \cdots \left(\frac{y_j}{j!}\right)^{r_j} \tag{5.14}$$

and

$$y_j = (U^{-1}\kappa)_j = \sum \frac{j!}{r_1! \cdots r_j!} \left(\frac{\kappa_1}{1!}\right)^{r_1} \cdots \left(\frac{\kappa_j}{j!}\right)^{r_j}, \tag{5.15}$$

$j = 1, \cdots, q$. Here, each summation extends over all the q-tuples $(r_1, \cdots, r_j) \in Z_+^j$ such that $\Sigma i r_i = j$. Further, $h = -1 + \Sigma r_i$.

5.5. Let us first consider the case $q = 2$, the case $q = 1$ being trivial. By $\kappa_1 = y_1$ and $\kappa_2 = y_2 - y_1^2$, the set $K[2]$ is defined by

$$0 \leqq \kappa_1 \leqq c, \qquad 0 \leqq \kappa_2 \leqq \kappa_1(c - \kappa_1). \tag{5.16}$$

It so happens that $K[2]$ is *convex*. Hence, the set $K[2]^n$ of all possible points $(\kappa_1(S_n), \kappa_2(S))$ coincides with $nK[2]$. Thus, $0 \leqq E(S_n) \leqq nc$ and

$$\text{Var}(S_n) \leqq E(S_n)\left[c - \frac{1}{n}E(S_n)\right], \tag{5.17}$$

and these inequalities cannot be improved.

5.6. Let us now turn to the case $q = 3$. The lower bound (5.10) on y_3 is of the form

$$y_2 = EZ^2 = E(Z^{1/2}\cdot Z^{3/2}) \leqq (y_1 y_3)^{1/2} \tag{5.18}$$

It is valid whenever $Z \geqq 0$ and is attained if and only if $\mu \in \mathscr{M}$ has a two-point support $\{0, \xi\}$ with $\xi \geqq 0$. In terms of cumulants this becomes

$$y_2^2 = (\kappa_1^2 + \kappa_2)^2 \leqq \kappa_1(\kappa_1^3 + 3\kappa_1\kappa_2 + \kappa_3); \tag{5.19}$$

thus,

$$\kappa_3 \geqq f(\kappa_1, \kappa_2) = (\kappa_2 - \kappa_1^2)\left(\frac{\kappa_2}{\kappa_1}\right). \tag{5.20}$$

The upper bound (5.10) on y_3 takes the form

$$\kappa_3 \leqq g(\kappa_1, \kappa_2) = \left((c - \kappa_1)^2 - \kappa_2\right)\left(\frac{\kappa_2}{1 - \kappa_1}\right), \tag{5.21}$$

(and is assumed if and only if $\mu \in \mathscr{M}$ has a two-point support $\{\xi, c\}$). We conclude that $K[3]$ may also be described as the set of points $(\kappa_1, \kappa_2, \kappa_3)$ in R^3 satisfying (5.16), (5.20), and (5.21).

LEMMA 5.2. *The $K[3]$ is not convex. Moreover, $K_3^* = \text{conv}\, K[3]$ is equal to the convex hull of the curve*

$$\Pi = \{\Pi_\rho = (\rho, \rho(c - \rho), \rho(c - \rho)(c - 2\rho)); 0 \leqq c\}. \tag{5.22}$$

It follows that $(z_1, z_2, z_3) \in K_3^*$ if and only if $(z_1, z_2) \in K[2]$, (that is, $0 \leqq z_2 \leqq z_1(1 - z_1)$) and, moreover,

$$-cz_2 + \frac{2z_2^2}{z_1} \leqq z_3 \leqq cz_2 - \frac{2z_2^2}{c - z_1}, \tag{5.23}$$

(if $0 < z_1 < c$; otherwise $z_2 = z_3 = 0$).

REMARK 5.1. The z_3 projection of K_3^* is nothing but $K[2] = K_2^*$. The z_1 projection is easily seen to be given by

$$0 \leqq z_2 \leqq \tfrac{1}{4}c^2, \qquad |z_3| \leqq z_2[c^2 - 4z_2]^{1/2}. \tag{5.24}$$

The z_2 projection is found to be

$$\begin{aligned} &-\tfrac{1}{8}c^2 z_1 \leqq z_3 \leqq z_1(c - z_1)(c - 2z_1), && \text{if} \quad 0 \leqq z_1 \leqq \tfrac{1}{4}c; \\ &-\tfrac{1}{8}c^2 z_1 \leqq z_3 \leqq \tfrac{1}{8}c^2(c - z_1), && \text{if} \quad \tfrac{1}{4}c \leqq z_1 \leqq \tfrac{3}{4}c; \\ &-z_1(c - z_1)(2z_1 - c) \leqq z_3 \leqq \tfrac{1}{8}c^2(c - z_1), && \text{if} \quad \tfrac{3}{4}c \leqq z_1 \leqq c. \end{aligned} \tag{5.25}$$

PROOF OF LEMMA 5.2. If $K[3]$ were convex, then f would be convex and g would be concave on the domain $K[2]$ defined by (5.16). On the contrary, consider the lower boundary of $K[3]$ on the cross section $\kappa_2 = (c - \rho)\kappa_1$, where $0 < \rho \leqq c$, $0 \leqq \kappa_1 \leqq \rho$. It is given by

$$\kappa_3 = f(\kappa_1, (c - \rho)\kappa_1) = (c - \rho)\kappa_1(c - \rho - \kappa_1). \tag{5.26}$$

The latter function is strictly concave (as soon as $\rho < c$) instead of convex. We also conclude that in forming the convex hull K_3^* of $K[3]$ the lower boundary of the cross section may as well be replaced by the pair of endpoints, namely, the point Π_0 corresponding to $\kappa_1 = 0$ and Π_ρ corresponding to $\kappa_1 = \rho$.

Similarly, the upper boundary of $K[3]$ in its cross section with $\kappa_2 = \rho(c - \kappa_1)$, $0 \leqq \rho \leqq c$, $\rho \leqq \kappa_1 \leqq c$, is given by the convex function

$$\kappa_3 = g(\kappa_1, \rho(c - \kappa_1)) = \rho(c - \kappa_1)(c - \kappa_1 - \rho). \tag{5.27}$$

In forming K_3^*, this part of the upper boundary may be replaced by the pair of endpoints Π_ρ and Π_c.

It follows that indeed K_3^* is precisely equal to the convex hull of the curve Π described in (5.22). The linear transformation

$$y_1 = z_1, \qquad y_2 = cz_1 - z_2, \qquad y_3 = c^2 z_1 - \tfrac{3}{2}z_2 + \tfrac{1}{2}z_3 \tag{5.28}$$

sends the point Π_ρ into a point y with coordinates $y_j = \rho^j$, $j = 1, 2, 3$. Thus, it sends K_3^* onto the corresponding convex hull which happens to be $M[3]$ (see (5.9)). The latter is determined by the inequalities $y_1^2 \leqq y_2 \leqq cy_1$ and (5.10). Transforming back, we conclude that K_3^* is determined by the inequalities $0 \leqq z_2 \leqq z_1(c - z_1)$ and (5.23).

THEOREM 5.1. *If S_n is a sum of n independent random variables $0 \leqq Z_i \leqq c$, then $z_j = (1/n)\kappa_j(S_n)$, $j = 1, 2, 3$, defines a point of K_3^*. Hence, we have, besides* (5.17), *that*

$$-c + \frac{2\operatorname{Var}(S_n)}{ES_n} \leqq \frac{\kappa_3(S_n)}{\operatorname{Var}(S_n)} \leqq c - \frac{2\operatorname{Var}(S_n)}{n - ES_n}. \tag{5.29}$$

Thus, $\kappa_3(S_n) > 0$ as soon as $\operatorname{Var}(S_n) > \tfrac{1}{2}c(ES_n)$. *Moreover, by* (5.24),

$$|\kappa_3(S_n)| \leqq \operatorname{Var}(S_n)\left[c^2 - \frac{4}{n}\operatorname{Var}(S_n)\right]^{1/2}. \tag{5.30}$$

The inequalities (5.29) *are sharp in the following sense. Let $z = (z_1, z_2, z_3)$ be a given point in* int (K_3^*). *Then for n sufficiently large there exists an admissible sum S_n with $(1/n)\kappa_j(S_n) = z_j$, $j = 1, 2, 3$.*

PROOF. Combine (5.6), Lemma 5.1, and Lemma 5.2.

REMARK 5.2. The point Π_ρ is realized by the cumulants of the measure μ with support $\{0, c\}$ and mass ρ/c at the point c. Thus, $\Pi \subset K[3]$; hence, $(1/n)\,\Pi^n \subset (1/n)\kappa[3]^n$ and each tends to K_3^* in the sense of Lemma 5.1.

It can be shown that to each point $z \in \operatorname{int}(K_3^*)$ there corresponds an integer n_0 such that for each integer $n \geqq n_0$ there exists a representation of z as $z =$

$\Sigma_{h=1}^{3} (n_h/n) \Pi_{\rho h}$ with $n_h \in Z_+$, $\Sigma_1^3 n_h = n$, $0 \leqq \rho_h \leqq c$, (all depending on n and z, $h = 1, 2, 3$). This implies that z can be realized as $z_j = (1/n)\kappa_j(S_n), j = 1, 2, 3$, by a sum of independent random variables $S_n = Z_1 + \cdots + Z_n$ such that each Z_i takes only the values 0 and c, while $Pr(Z_i = c) = \rho_h/c$ for exactly n_h indices $i = 1, \cdots, n, h = 1, 2, 3$.

5.7. Let us finally consider the case $q = 4$, restricting ourselves to a lower bound on $\kappa_4(S_n)$.

If Z is any random variable with a finite fourth moment, then the best lower bound for the cumulant $\kappa_4 = \kappa_4(Z)$ in terms of the lower cumulants is given by

$$\kappa_4 \geqq \frac{\kappa_3^2}{\kappa_2} - 2\kappa_2^2. \tag{5.31}$$

If $0 \leqq Z \leqq c$, then (5.31) follows immediately from the first inequality (5.11). Since the validity of (5.31) is not affected by a change of location or scale, it follows that (5.31) holds for each bounded Z and thus for each Z with a finite fourth moment.

The argument further shows that $K[4]$ can be defined by the inequalities (5.16), (5.20), (5.21), (5.31), and a somewhat more complicated upper bound on κ_4 which may be derived from the second inequality (5.11) (and which will involve c).

LEMMA 5.3. *The lower boundary of $K_4^* =$ conv $K[4]$ is the same as the lower boundary of the convex hull of the curve $\Sigma = \{\Sigma_\rho; 0 \leqq \rho \leqq c\}$, where*

$$\Sigma_\rho = \big(\rho, \rho(c - \rho), \rho(c - \rho)(c - 2\rho), \rho(c - \rho)(c^2 - 6c\rho + 6\rho^2)\big). \tag{5.32}$$

Moreover, $z \in R^4$ belongs to K_4^* if and only if:

(i) $(z_1, z_2, z_3) \in K_3^*$; *that is,* $0 \leqq z_2 \leqq z_1(1 - z_1)$ *and* (5.23) *hold;*

(ii) *we have the lower bound*

$$z_4 \geqq \frac{3z_3^2}{2z_2} - \frac{1}{2}c^2 z_2; \tag{5.33}$$

(iii) z_4 *satisfies an analogous upper bound* $z_4 \leqq h(z_1, z_2, z_3)$ *which will not be specified.*

REMARK 5.3. We shall need to relate the inequalities (5.33) and

$$z_4 \geqq \frac{z_3^2}{z_2} - 2z_2^2, \tag{5.34}$$

where $z_2 > 0$. In fact, (5.34) implies (5.33) precisely when

$$\frac{z_3^2}{z_2} - 2z_2^2 \geqq \frac{3}{2}\frac{z_3^2}{z_2} - \tfrac{1}{2}c^2 z_2, \tag{5.35}$$

which is equivalent to the pair of inequalities

$$z_2 \leqq \tfrac{1}{4}c^2, \qquad |z_3| \leqq z_2[c^2 - 4z_2]^{1/2}. \tag{5.36}$$

By (5.24) this is true if and only if (z_2, z_3) corresponds to some point $z = (z_1, z_2, z_2) \in K_3^*$. In all other cases, in particular when $z_2 > \frac{1}{4}c^2$, we have that (5.33) implies (5.34).

PROOF OF LEMMA 5.3. The vertical projection (in the κ_4 direction) of $K[4]$ is precisely $K[3]$. Hence, the vertical projection of conv $K[4] = K_4^*$ is precisely conv $K[3] = K_3^*$ and the latter is completely described by Lemma 5.2.

Consider the curve Σ described by (5.32). We easily verify that the point Σ_ρ of the curve is realized by the first four cumulants of the probability measure μ_ρ having the two-point support $\{0, c\}$ and a mass ρ/c at c. Hence, $\Sigma \subset K[4]$ and therefore, conv $(\Sigma) \subset$ conv $K[4] = K_4^*$. By Lemma 5.2, the vertical projection of conv (Σ) is precisely conv $(\Pi) = K_3^*$.

For each point $z = (z_1, z_2, z_3)$ in K_3^*, let us define

$$\phi(z) = \inf \{\zeta: (z_1, z_2, z_3, \zeta) \in K_4^*\} \tag{5.37}$$

and

$$\psi(z) = \inf \{\zeta: (z_1, z_2, z_3, \zeta) \in \operatorname{conv}(\Sigma)\}. \tag{5.38}$$

Clearly, both ϕ and ψ are convex functions on K_3^*. Further, $\phi(z) \leqq \psi(z)$ since conv $(\Sigma) \subset K_4^*$. We shall prove below that

$$\psi(z) = \frac{3}{2}\frac{z_3^2}{z_2} - \frac{1}{2}c^2 z_2 \quad \text{for all } z \in K_3^*, \tag{5.39}$$

and it would suffice to prove that $\phi \equiv \psi$.

In fact, we know from (5.31) that, for each $z \in K[4]$, $z_4 \geqq z_3^2/z_2 - 2z_2^2$. Using formula (5.39) and Remark 5.3, we conclude that $z_4 \geqq \psi(z_1, z_2, z_3)$ for each $z \in K[4]$, and hence, for each $z \in$ conv $K[4] = K_4^*$ since the function ψ is convex. This in turn implies that $\phi(z) \geqq \psi(z)$, and hence, $\phi(z) = \psi(z)$ for all $z \in K_3^*$.

It only remains to verify the formula (5.39) for the function ψ defined by (5.38). One proof would be to derive it from the second inequality (5.11) by a transformation analogous to the one used in the last part of the proof of Lemma 5.2. Another proof would be as follows.

First, introduce $\tilde{\psi}(z) = \frac{3}{2}z_3^2/z_2 - \frac{1}{2}c^2 z_2$. Then $\tilde{\psi}$ is convex throughout the region $z_2 > 0$, as can for instance be seen from the formula

$$\tilde{\psi}(z) = \sup_u \left[-\tfrac{1}{2}(3u^2 + c^2)z_2 + 3uz_3\right]. \tag{5.40}$$

It follows that the region $W = \{z \in R^4: z' \in K_3^*, z_4 \geqq \tilde{\psi}(z')\}$ is convex; $z' = (z_1, z_2, z_3)$ if $z = (z_1, z_2, z_3, z_4)$.

Second, an easy computation shows that

$$z_4 = \tilde{\psi}(z') \text{ for each } z \in \Sigma; \tag{5.41}$$

$\tilde{\psi}(\Sigma_\rho) = 0$ if $\rho = 0$ or $\rho = c$. Hence, $\Sigma \subset W$, and hence, conv $(\Sigma) \subset W$; thus, $\psi(z) \geqq \tilde{\psi}(z)$ throughout K_3^*.

Third, if $z \in \Sigma$ then, using (5.38) and (5.41), we have that $\psi(z') \leqq z_4 = \tilde{\psi}(z')$; thus, $\psi(z') = \tilde{\psi}(z')$.

Fourth, on the line segment $z_3 = \lambda z_2$, $|\lambda| < c$, $z \in K_3^*$, the function $\tilde{\psi}(z)$ is linear in z_2 while $\psi(z)$ is convex so that the difference $\chi(z) = \psi(z) - \tilde{\psi}(z)$ is convex and nonnegative. Moreover, $\chi(z) = 0$ at both end points (which correspond to points of Σ), and hence, throughout the entire segment. This proves that $\psi(z) = \tilde{\psi}(z)$ throughout K_3^* and establishes (5.39).

THEOREM 5.2. *Suppose $S_n = Z_1 + \cdots + Z_n$ is the sum of n independent random variables $0 \leqq Z_i \leqq c$. Then*

$$\kappa_4(S_n) \geqq \frac{3}{2}\frac{\kappa_3(S_n)^2}{\kappa_2(S_n)} - \tfrac{1}{2}c^2\kappa_2(S_n). \tag{5.42}$$

The inequality (5.42) is sharp in the sense that given $z \in \operatorname{int}(K_4^*)$ (in particular $z_4 > \frac{3}{2}z_3^2/z_2 - \frac{1}{2}c^2z_2$) there exists for each sufficiently large integer n an admissible sum S_n with $(1/n)\kappa_j(S_n) = z_j$, $j = 1, 2, 3, 4$.

PROOF. Combine (5.6), Lemma 5.1, and Lemma 5.3.

REMARK 5.4. In view of (5.31), we also have

$$\kappa_4(S_n) \geqq \frac{\kappa_3(S_n)^2}{\kappa_2(S_n)} - 2\kappa_2(S_n)^2. \tag{5.43}$$

As follows from Remark 5.3, (5.43) is better than (5.42) if and only if $(\kappa_2(S_n, \kappa_3(S_n))$ corresponds to a point $(z_1, \kappa_2(S_n), \kappa_3(S_n))$ of $K_3^* = \operatorname{conv} K[3]$, for some choice of z_1. For n large this is rarely the case. If $\kappa_2(S_n)$ is large, then (5.42) is obviously much more precise than (5.43).

6. Exponential bounds

6.1. In this section we shall again be interested in a sum $S_n = Z_1 + \cdots + Z_n$ of independent real valued random variables. We shall make two assumptions.

(i) Let $a < b$ be given finite constants and assume that $a \leqq Z_i \leqq b$ for all $i = 1, \cdots, n$.

(ii) Let m be a given positive integer and $c_1, \cdots, c_m$ given numbers. We assume that

$$E(Z_i^j) = c_j \quad \text{for } j = 1, \cdots, m \text{ and all } i = 1, \cdots, n. \tag{6.1}$$

Naturally, $c_1, \cdots, c_m$ must be such that there do exist such random variables $a \leqq Z_i \leqq b$; that is, $c = (c_1, \cdots, c_m)$ must be a point of the corresponding moment space $M[m]$.

Most of the difficulties encountered in Section 5 concerning the precise relations between the first q moments of S_n had to do with the fact that the transformation (5.14) between moments and cumulants is a nonlinear transformation which may transform a convex set $M[q]$ into a nonconvex set $K[q]$.

In the present section, we want to exploit the fact that κ_j happens to be linear in the higher moments y_j. Fixing the lower moments as in (6.1) will make the

cumulants $\kappa_j(Z_i)$ with $j \leqq 2m + 1$ into an affine function of the unknown moments $y_r(Z_i) = E(Z_i^r)$. More precisely, (5.14) and (6.1) together imply a relation of the form

$$\kappa_j(Z_i) = a_j + \sum_{r=m+1}^{j} b_{j,r} y_r(Z_i), \qquad j = m + 1, \cdots, 2m + 1, \tag{6.2}$$

valid for all $i = 1, \cdots, n$. Here, the coefficients a_j and $b_{j,r}$ are known constants (depending only on $c_1, \cdots, c_m$).

The collection M of possible moment points $\bar{y} = (y_{m+1}(Z_i), \cdots, y_{2m+1}(Z_i))$ is a known convex subset of R^{m+1}. Hence, from (6.2), also the collection K of all possible cumulant points $\bar{z} = (\kappa_{m+1}(Z_i), \cdots, \kappa_{2m+1}(Z_i))$ is a known *convex* subset K of R^{m+1}, the same set for all $i = 1, \cdots, n$. By (5.3), the set of all possible cumulant points $(\kappa_{m+1}(S_n), \cdots, \kappa_{2m+1}(S_n))$ of S_n is precisely the set $K + K + \cdots + K = K^n = nK$, the latter because K is convex. This leads to the statement

$$\left(\frac{1}{n}\kappa_{m+1}(S_n), \cdots, \frac{1}{n}\kappa_{2m+1}(S_n)\right) \in K. \tag{6.3}$$

And for each integer n this is more or less all we can say about these cumulants, that is, about the moments $E(S_n^j)$ with $j \leqq 2m + 1$; the transformation (6.2) cannot be used, however, for Z_i replaced by S_n, but has to be modified since S_n has its lower moments different from Z_i.

Suppose $Z_1, \cdots, Z_n$ besides satisfying (i), (ii) also satisfy:

(iii) the Z_i are identically distributed.

Then

$$\left(\frac{1}{n}\kappa_{m+1}(S_n), \cdots, \frac{1}{n}\kappa_{2m+1}(S_n)\right) = (\kappa_{m+1}(Z_1), \cdots, \kappa_{2m+1}(Z_1)), \tag{6.4}$$

but still the latter can be any point in K. In other words, the relation (6.3) cannot be improved at all. That is, any relation between the moments $E(S_n^j)$ with $1 \leqq j \leqq 2m + 1$ which is universally true under the assumptions (i), (ii), and (iii) is thus also universally true under the assumptions (i) and (ii) alone.

6.2. Let us now consider the following related problem. Namely, we still assume (i) and (ii), but we now want to add the assumption:

(iv) let q be a given positive integer with $m \leqq q \leqq 2m + 1$ and assume that

$$E(S_n^j) = d_j \quad \text{for all } m < j \leqq q, \tag{6.5}$$

where the d_j are given numbers. Naturally (6.5) is void when $q = m$. Since (i), (ii), and (iv) must have a common solution, the d_j cannot be entirely arbitrary. Note that the distributions of the Z_i will be different in general.

We shall show that in this situation it is not difficult to obtain an *exact* formula (in terms of a, b, the c_j and d_j) for the quantity

$$A(t) = \sup_{S_n} \log E(\exp\{tS_n\}), \tag{6.6}$$

where S_n ranges through all sums of the above type, while t denotes a fixed

constant, $t \neq 0$. Such an exact formula may be useful in connection with the well-known inequality

$$Pr(S_n \geqq nx) \leqq \exp\{-tnx\}\, E(\exp\{tS_n\}) \leqq \exp\{-tnx + A(t)\} \quad \text{if } t > 0. \tag{6.7}$$

6.3. Let us first reformulate the above problem. In the first place, we may rewrite (6.6) as

$$A(t) = \sup_{Z_1,\cdots,Z_n} \sum_{i=1}^{n} \log E(\exp\{tZ_i\}), \tag{6.8}$$

where $(Z_1, \cdots, Z_n)$ ranges through the n-tuples of random variables satisfying (i), (ii), while further $\Sigma_{i=1}^{n} \kappa_j(Z_i)$ is equal to a given number when $m < j \leqq q$. Using (6.1), (6.2), and $q \leqq 2m + 1$, the latter condition can be rewritten as

$$\sum_{i=1}^{n} E(Z_i^j) = e_j \quad \text{for all } m < j \leqq q, \tag{6.9}$$

where the e_j denote given numbers, which are easily calculated from the c_j and d_j. We now have that in (6.8) the Z_i range through the n-tuples satisfying (i), (ii) and (6.9); in the present formulation the independence of the Z_i is no longer important. Let us now proceed to show that *the supremum in* (6.8) *is attained for the case where the* Z_i *are identically distributed*. This will reduce our problem to the more or less classical one of finding

$$A(t) = n \sup \log E(\exp\{tZ\}), \tag{6.10}$$

where Z is a random variable subject only to the conditions that $a \leqq Z \leqq b$ and that $E(Z^j) = c_j$, $j = 1, \cdots, q$, with $c_1, \cdots, c_q$ as given numbers.

6.4. The above problem, to determine the maximum (6.8) subject to the conditions $a \leqq Z_i \leqq b$, (6.1), and (6.9), may be generalized as follows.

Namely, consider a measurable space X and a given convex class $\mathcal{M}$ of probability measures on X. Let further ψ and g_j, $j = 1, \cdots, r$, be given real valued measurable functions on X which are integrable relative to each $\mu \in \mathcal{M}$. Finally, let ϕ be a given real valued and *concave* function defined on an interval containing the range of ψ. Our problem will be to determine

$$\max \sum_{i=1}^{n} \phi[E(\psi(Z_i))]. \tag{6.11}$$

Here, $Z_1, \cdots, Z_n$ will denote random variables taking values in X such that (a) the distribution μ_i of Z_i belongs to the given class $\mathcal{M}$, $i = 1, \cdots, n$; (b) the Z_i must further satisfy the side conditions

$$\sum_{i=1}^{n} E(g_j(Z_i)) = e_j \quad \text{for } j = 1, \cdots, r. \tag{6.12}$$

Here, the e_j denote given numbers such that there do exist n-tuples of random variables satisfying the stated properties.

6.5. In our original problem we have $X = [a, b]$. Further, $\mathcal{M}$ may be taken as the class of all probability measures on X satisfying $\int x^j \mu(dx) = c_j$, $j = 1, \cdots, m$. Moreover, in (6.12) we take $r = q - m$ and $g_j(x) = x^{m+j}$. Finally, $\phi(u) = \log u$ and $\psi(x) = e^{tx}$. Observe that $\log u$ is indeed concave on the range $(0, \infty)$ of ψ.

LEMMA 6.1. *In the above general problem of determining* (6.11), *the supremum is not decreased by adding the additional condition that $Z_1, \cdots, Z_n$ all have the same distribution.*

PROOF. Let Z_i have distribution $\mu_i \in \mathcal{M}$, $i = 1, \cdots, n$, and put $\mu = (\mu_1 + \cdots + \mu_n)/n$. Then $\mu \in \mathcal{M}$ since $\mathcal{M}$ is convex. It suffices to prove that

$$\frac{1}{n}\sum_{i=1}^{n} \phi\left(\int \psi(x)\mu_i(dx)\right) \leqq \phi\left(\int \psi(x)\mu(dx)\right). \tag{6.13}$$

But this follows immediately from Jensen's inequality.

6.6. Let us return to the original problem (6.6). Applying Lemma 6.1, we conclude that (6.8) may be reformulated as in (6.10). Now observe that the $q + 2$ functions $g_j(x) = x^j$, $j = 0, 1, \cdots, q$, and $g_{q+1}(x) = e^{tx}$ together form a Chebyshev system over every subinterval of R, (see [7] p. 45, [4] pp. 6 and 376). It follows from a classical result due to Markov (see [7], p. 61, [4], pp. 55 and 80) that the supremum (6.10) is attained for either the upper or the lower principal representation (depending on the sign of t) corresponding to the preassigned moment point $y = (c_1, \cdots, c_q)$, (provided $y \in \operatorname{int}(M[q])$ which we shall assume).

Here, by a principal representation we mean a probability measure on $[a, b]$ having the preassigned moments $c_1, \cdots, c_q$ and further a finite support S consisting of $\frac{1}{2}(q + 1)$ points (an end point a or b counting only as half a point). There are exactly two such principal representations of y. Observe that the pair of principal representations is totally independent of t. Assuming that $t > 0$, the principal representation maximizing $\int e^{tx}\mu(dx)$ always has the right end point b in its support and this requirement uniquely determines the principal representation needed; if $t < 0$ then the left end point a must be in the support.

THEOREM 6.1. *Let $S_n = Z_1 + \cdots + Z_n$, where $Z_1, \cdots, Z_n$ are independent random variables with possibly different distributions such that $|Z_i| \leqq 1$ and $E(Z_i) = 0$, $i = 1, \cdots, n$. Put* $\operatorname{Var}(S_n) = s^2 = nc$, $0 \leqq c \leqq 1$, *and let $t > 0$ be a given constant. Then*

$$E(e^{tS_n}) \leqq \left[\frac{c}{1 + c}e^t + \frac{1}{1 + c}e^{-ct}\right]^n, \tag{6.14}$$

and this inequality cannot be improved.

PROOF. We must compute (6.6) subject to $|Z_i| \leqq 1$, $E(Z_i) = 0$ and $E(S_n^2) = nc$. This is a very special case of the general problem (6.6). We already showed that we may assume the Z_i to be identically distributed; hence, we must prove that

$$E(e^{tZ}) \leqq \frac{c}{1 + c}e^t + \frac{1}{1 + c}e^{-ct}, \tag{6.15}$$

when it is known that $|Z| \leqq 1$, $E(Z) = 0$, $E(Z^2) = s^2/n = c$. The largest possible

value $E(e^{tZ})$ must be attained by the principal representation corresponding to the moments $y_1 = 0$, $y_2 = c$ with a support S of size $\frac{3}{2}$ and such that $1 \in S$; we have $(y_1, y_2) \in \text{int } M[2]$ provided $0 < c < 1$; the relation (6.15) is obvious when $c = 0$ or $c = 1$. It follows that necessarily $S = \{-c, 1\}$, while $Pr(Z = 1) = c/(1 + c)$. This proves (6.15). If each Z_i has the latter distribution, then (6.15) holds with the equality sign and hence cannot be improved.

6.7. As a final application, suppose $S_n = Z_1 + \cdots + Z_n$ is a sum of independent random variables such that $a \leqq Z_i \leqq b$ and $E(Z_i) = 0, i = 1, \cdots, n$; here, $a < 0 < b$ are fixed. We want to establish the best possible upper bound on $E(\exp\{tS_n\})$, $t > 0$, in terms of $E(S_n^2) = s^2 = nc$ and $E(S_n^3) = \rho^3 = nd$, say. There does exist a random variable Z with $a \leqq Z \leqq b$ and $E(Z) = 0, EZ^2 = c$, $EZ^3 = d$ (see (6.3)). Moreover, from (6.10) and Section 6.6,

$$\sup_{S_n} E(\exp\{tS_n\}) = (E \exp\{tZ\})^n, \tag{6.16}$$

where $a \leqq Z \leqq b$ has as its distribution the principal representation of the set of moments $y_1 = 0, y_2 = c, y_3 = d$ and with b in the support S; here, we assume that y is interior. Further, the support has $\frac{1}{2}(3 + 1) = 2$ points (counting end points half), implying that it is of the form $S = \{a, \xi, b\}$ with $a < \xi < b$. This leads to the equations $pa^j + q\xi^j + rb^j = y_j$, $j = 0, 1, 2, 3$, $y_0 = 1$, $y_1 = 0$, $y_2 = c, y_3 = d$, so that $\xi = (d - ac - bc)/(ab + c)$. The desired *optimal* upper bound is given by

$$E(\exp\{tS_n\}) \leqq (pe^{at} + qe^{\xi t} + re^{bt})^n, \tag{6.17}$$

which is easily computed. If desired, we can also derive a bound

$$Pr(S_n \geqq nx) \leqq [e^{-xt}(pe^{at} + qe^{\xi t} + re^{bt})]^n, \tag{6.18}$$

provided $t > 0$, $a < x < b$. The best value of t is obtained by putting the logarithmic derivative with respect to t equal to 0. In the special case $\xi = \frac{1}{2}(a + b)$, this leads to a simple quadratic equation; this happens when $d = (a + b)c, d = 0$ if $a = -b$.

REFERENCES

[1] N. Bourbaki, *Eléments de Mathématique, Livre* VI, *Intégration* Ch. 1–4, Paris, Hermann, 1952.

[2] W. R. Emerson and F. P. Greenleaf, "Asymptotic behavior of products $C^p = C + \cdots + C$ in locally compact abelian groups," *Trans. Amer. Math. Soc.*, Vol. 145 (1969), pp. 171–204.

[3] Ky Fan, *Convex Sets and Their Applications*, Argonne National Laboratory, Applied Mathematics Division, Summer Lectures, 1959.

[4] S. J. Karlin and W. J. Studden, *Tchebycheff Systems: with Applications in Analysis and Statistics*, New York, Interscience, 1966.

[5] J. H. B. Kemperman, "On the sharpness of Tchebycheff type inequalities," *Indagationes Mathematicae*, Vol. 27 (1965), pp. 554–601.

[6] ———, "The general moment problem, a geometric approach," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 93–122.

[7] M. G. Krein, "The ideas of P. L. Čebyšev and A. A. Markov in the theory of limiting values of integrals, and their further developments," *Uspehi Mat. Nauk* (N. S.), Vol. 6 (1951), pp. 3–120. (In Russian; English translation: *Amer. Math. Soc. Translations*, Series 2, Vol. 12 (1959), pp. 1–121.)